Theoretical Computer Science 54 (1987) 299-313
North-Holland

CANONICAL FORMS FOR INTERVAL FUNCTIONS

Seymour GINSBURG*

Computer Science Department, University of Southern California, Los Angeles, CA 90089, U.S.A.

Chang-jie TANG**

Computer Science Department, Sichuan University, Chengdu, Sichuan, China

Communicated by M. Nivat
Received October 1986

Abstract. The following six types of functions are considered which, when applied to a word, return an interval of that word: (1) prefix_k, which returns the prefix of length k; (2) suffix_k, which returns the suffix of length k; (3) chop^k, which returns all but the first k symbols; (4) chop_k, which returns all but the last k symbols; (5) Fprefix_λ, which returns the first λ part of the word; and (6) Fsuffix_λ, which returns the last λ part of the word. (Here k is a nonnegative integer and λ a real number, $0<\lambda<1$.) The main result is that any composition of functions of the first four types which is not a trivial function is represented in exactly one of the forms $\text{prefix}_j\text{chop}^l\text{chop}_m$, $\text{suffix}_k\text{chop}^l\text{chop}_m$, $\text{chop}^l\text{chop}_m$, or $\text{prefix}_2\text{suffix}_1$, and is unique within that form. A discussion is also given about canonical forms for the composition of functions of all types.

Introduction

In [1] a record-based, computation-oriented data model was introduced for describing historical data (called 'object history' and represented by a sequence of 'computation tuples'). In a subsequent paper [2], object histories were studied with respect to a variety of 'interval queries'. The present investigation began by observing that some of the interval queries of [2] were independent of the content of the computation tuples, i.e., were dependent only on properties of sequences, and thus were really mappings on words, i.e., were interval functions. We selected three basic types of these interval functions, added a fourth which was symmetric to one of the three, and studied the family of functions obtained by composition. Our main result, Theorem 1.1, asserts that each interval function in the resulting family has a unique[1] canonical representation, namely each such function is uniquely expressed as the composition (in a certain manner) of at most three members of the four types. While originating from a database source, this theorem is not a database result. Rather, it is a purely mathematical statement about a special class of interval functions.

* This author was supported in part by the National Science Foundation under Grant DCR-8318752.
** Research done while on leave from Sichuan University, Chengdu, Sichuan, China.
[1] If the interval function is trivial, uniqueness need not hold.

Consider the following general problem: 'Let H be a set of expressions, each representing a function, and let H^+ be the set of expressions $\{h_1 \dots h_n \mid n \geq 1$, each h_i in $H\}$. Then H^+ may be viewed as a set of functions in which concatenation is composition. Does there exist an integer t and a subset H' of H^+ such that $H' = H^+$ (as sets of functions) and each h in H' is the composition of at most t functions in H?' When the answer is yes, we shall refer to H' as a set of 'canonical forms' or 'canonical functions' (for H^+). All the results of this paper concern the existence of t and H' for certain sets H. Indeed, our presentation is divided into three sections. Section 1 defines the four basic types of interval functions mentioned earlier and states the main theorem. (Obviously, the main theorem provides an affirmative answer, with $t = 3$, for the case when H consists of the four basic types of functions.) Section 2 gives the proof, in the course of which an effective procedure for determining the canonical form is exhibited. Also in Section 2 is the result that, for the set H consisting of the four basic types of functions and for $t \leq 2$, there is no H'. The last section introduces interval functions which return a fraction (e.g., first half or last third) of the original interval, and considers canonical forms for the composition of these new functions either alone or in conjunction with the four basic types. In particular, it is shown that no set of canonical forms exists.

1. A family of interval functions

In this section, we motivate and then formalize the main functions in the paper. We conclude with a statement of the principal theorem.

Suppose that John Smith opens a checking account (interest posted monthly) on 2 January 1984. On 5 January, the checking-account history is as given in Table 1. The following four kinds of queries arise naturally:

(Q1) Retrieve the first k tuples of Smith's checking-account history.

(Q2) Retrieve the last k tuples of Smith's checking-account history.

(Q3) Retrieve all of Smith's checking-account history except the first k transactions.

(Q4) Retrieve all of Smith's checking-account history except the last k transactions.

Table 1.

	Date	Action	Amount	Balance
a_1	2 Jan. 1984	Deposit	500	500
a_2	3 Jan. 1984	Deposit	500	1000
a_3	4 Jan. 1984	Deposit	1000	2000
a_4	4 Jan. 1984	Check	500	1500
a_5	5 Jan. 1984	Check	200	1300
a_6	5 Jan. 1984	Check	500	800

(Types (Q1), (Q2) and (Q3) are explicitly mentioned in [2], while (Q4), symmetric to (Q3), is not.) Note that the answers to queries of any of the above types

(1) are intervals of the original history, and
(2) depend on the sequential position of the tuples and not on the contents of the tuples.

We shall abstract the queries of types (Q1)-(Q4) from sequences of tuples to sequences of symbols, and study canonical forms (as defined in the Introduction) for the family obtained by composition of queries of these types.

Formally, Σ henceforth is a (possibly infinite) set of at least two elements. The symbols a, b, c and d, possibly subscripted, always denote elements of Σ. A *word* over Σ is a finite sequence $a_1 \dots a_m$ of elements of Σ. The empty word ($m = 0$) is denoted by ε. The set of all words over Σ is denoted by Σ^*.

We now define the first two of our four basic types of interval queries, henceforth referred to as interval functions. (They are called interval functions because, when defined, they map a word w into an interval of[2] w.)

Notation. For each nonnegative integer k, let prefix_k and suffix_k be the (partial) functions from Σ^* to Σ^* defined for each word $a_1 \dots a_m$ by $\text{prefix}_k(a_1 \dots a_m) = a_1 \dots a_k$ if $m \geq k \geq 0$ and $\emptyset$ (i.e., undefined) otherwise, and $\text{suffix}_k(a_1 \dots a_m) = a_{m-k+1} \dots a_m$ if $m \geq k \geq 0$ and $\emptyset$ otherwise.

Thus, prefix_k applied to w returns the prefix of w of length[3] k if $|w| \geq k$. Similarly, suffix_k applied to w returns the suffix of w of length k if $|w| \geq k$. Clearly, prefix_k and suffix_k correspond to (Q1) and (Q2) respectively.

Notation. Let E (the *trivial* mapping) be the function from Σ^* to Σ^* defined by $E(w) = \varepsilon$ for all w in Σ^*. Let Φ be the (partial) function from Σ^* to Σ^* defined by $\Phi(w) = \emptyset$ for all w in Σ^*.

Note that $\text{prefix}_0 = \text{suffix}_0 = E$, and[4] $\text{prefix}_{k+1}\text{suffix}_k = \text{suffix}_{k+1}\text{prefix}_k = \Phi$ for all $k \geq 0$. Also, $\Phi f = f\Phi = \Phi$ for all (partial) functions f from Σ^* to Σ^*.

We now present the two remaining types of basic interval functions.

Notation. For each nonnegative integer k, let chop^k and chop_k be the (partial) functions from Σ^* to Σ^* defined for each $a_1 \dots a_m$ by $\text{chop}^k(a_1 \dots a_m) = a_{k+1} \dots a_m$ and $\text{chop}_k(a_1 \dots a_m) = a_1 \dots a_{m-k}$ if $m \geq k \geq 0$, and is Φ otherwise.

Clearly, chop^k and chop_k correspond to (Q3) and (Q4) respectively. Also, $\text{chop}^0(w) = \text{chop}_0(w) = w$ for each w in Σ^*.

[2] An *interval* of a word $a_1 \dots a_m$ is any word of the form $a_i a_{i+1} \dots a_j$, $i - 1 \leq j \leq m$.

[3] For each word $w = a_1 \dots a_m$ in Σ^+, the *length* of $a_1 \dots a_m$, denoted $|w|$, is m. The *length* of ε, denoted $|\varepsilon|$, is 0.

[4] Let f_1 and f_2 be (partial) functions X to X. Then $f_1 f_2$, the *composition* of f_2 and f_1, is the (partial) mapping from X to X defined by $f_1 f_2(x) = f_1(f_2(x))$ for all x in X.

As already mentioned, we are interested in canonical forms for the family of all functions obtained by repeated composition of functions of any of the four basic types. Accordingly, we have the following symbolism.

Notation. Let

$$F=\{\text{prefix}_j, \text{suffix}_k, \text{chop}^l, \text{chop}_m \mid j, k, l, m \text{ nonnegative integers}\}$$

and

$$F^+=\{f_1 \ldots f_n \mid f_i \text{ in } F, 1 \leqslant i \leqslant n\}.$$

Obviously, each function in F^+ is an interval mapping. Also, each function f in F (thus in F^+) has the property, described informally in (2), that if $f(a_1 \ldots a_n)= a_i \ldots a_j$ for some $a_1 \ldots a_n$, i and j, then $f(b_1 \ldots b_n)=b_i \ldots b_j$ for all words $b_1 \ldots b_n$.

Using the previous notation, we are now ready to state our main result.

Theorem 1.1 (Main Theorem). *Each f in F^+ is expressible in the form*

(i) $\text{prefix}_j\text{chop}^l\text{chop}_m$,
(ii) $\text{suffix}_k\text{chop}^l\text{chop}_m$,
(iii) $\text{chop}^l\text{chop}_m$, *or*
(iv) $\text{prefix}_2\text{suffix}_1$ $(=\Phi)$.

In addition, if $f \neq E$, then f is in exactly one of the above forms, is uniquely expressible within that form,[5] *and if* (i) (*resp.* (ii)) *then j* (*resp. k*) *is positive.*

The proof is quite detailed and will consume most of the next section.

Note that the last phrase in the last sentence of the statement of Theorem 1.1 follows automatically from the fact that $f \neq E$.

In accordance with the terminology mentioned in the Introduction, we have the following definition.

Definition. Each expression of type (i), (ii), (iii) or (iv) of Theorem 1.1 is called a *canonical* form (for f in F^+).

Example 1.2. Let $f=\text{chop}^2\,\text{prefix}_5\,\text{chop}^2\,\text{suffix}_{10}\,\text{prefix}_{17}\,\text{chop}_8$ and $f'=\text{prefix}_3\,\text{chop}^{11}\,\text{chop}_{11}$. The expression f is not in canonical form, whereas f' is. Furthermore, $f=f'$! Indeed, for each word $w=a_1 \ldots a_n$, $f(w)=\emptyset=f'(w)$ if $n<25$, and $f(w)=a_{12}a_{13}a_{14}= f'(w)$ if $n \geqslant 25$.

[5] Thus, if $f=\text{prefix}_j\text{chop}^l\text{chop}_m$, f in $F^+-\{E\}$, for some j, l and m, then f cannot be expressed in either form (ii), (iii) or (iv). Moreover, if $f=\text{prefix}_{j'}f\text{chop}^{l'}\text{chop}_{m'}$, then $j=j'$, $l=l'$ and $m=m'$.

2. Proof of the main theorem

In this section we establish the proof of Theorem 1.1. In the course of our argument, we present a procedure for obtaining the canonical form (Theorem 2.3).

We start with a preliminary result which expresses the composition of various pairs of our basic types in terms of either the composition of other pairs or in terms of one type.

Lemma 2.1. *Let j and k be nonnegative integers. Then*

(a1) $\text{chop}_j\text{prefix}_k = \begin{cases} \text{prefix}_{k-j}\text{chop}_j & \textit{if } j \leqslant k, \\ \Phi & \textit{otherwise}; \end{cases}$

(a2) $\text{chop}_j\text{suffix}_k = \begin{cases} \text{suffix}_{k-j}\text{chop}_j & \textit{if } j \leqslant k, \\ \Phi & \textit{otherwise}; \end{cases}$

(a3) $\text{chop}_j\text{chop}^k = \text{chop}^k\text{chop}_j$

(a4) $\text{chop}_j\text{chop}_k = \text{chop}_{j+k}$

(b1) $\text{chop}^j\text{suffix}_k = \begin{cases} \text{suffix}_{k-j}\text{chop}^j & \textit{if } j \leqslant k, \\ \Phi & \textit{otherwise}; \end{cases}$

(b2) $\text{chop}^j\text{prefix}_k = \begin{cases} \text{prefix}_{k-j}\text{chop}^j & \textit{if } j \leqslant k, \\ \Phi & \textit{otherwise}; \end{cases}$

(b3) $\text{chop}^j\text{chop}^k = \text{chop}^{j+k}$

(c1) $\text{prefix}_j\text{prefix}_k = \begin{cases} \text{prefix}_j\text{chop}_{k-j} & \textit{if } j \leqslant k, \\ \Phi & \textit{otherwise}; \end{cases}$

(c2) $\text{prefix}_j\text{suffix}_k = \begin{cases} \text{suffix}_j\text{chop}_{k-j} & \textit{if } j \leqslant k, \\ \Phi & \textit{otherwise}; \end{cases}$

(d1) $\text{suffix}_j\text{suffix}_k = \begin{cases} \text{suffix}_j\text{chop}^{k-j} & \textit{if } j \leqslant k, \\ \Phi & \textit{otherwise}; \end{cases}$

(d2) $\text{suffix}_j\text{prefix}_k = \begin{cases} \text{prefix}_j\text{chop}^{k-j} & \textit{if } j \leqslant k, \\ \Phi & \textit{otherwise}. \end{cases}$

Proof. Clearly (a3), (a4) and (b3) are obvious; (b1) is similar to (a1); (b2) is similar to (a2); (d1) is similar to (c1) and (d2) is similar to (c2). It thus suffices to examine (a1), (a2), (c1), and (c2). Let $w = a_1 \ldots a_n$.

Consider (a1). Suppose $j \leqslant k$. If $k \leqslant n$, then

$$\text{chop}_j\text{prefix}_k(w) = a_1 \ldots a_{k-j} = \text{prefix}_{k-j}\text{chop}_j(w),$$

and if $n < k$, then $\text{chop}_j\text{prefix}_k(w) = \emptyset = \text{prefix}_{k-j}\text{chop}_j(w)$. Suppose $k < j$. Clearly, $\text{chop}_j\text{prefix}_k(w) = \emptyset = \Phi(w)$.

Consider (a2). Suppose $j \leqslant k$. If $k \leqslant n$, then

$$\text{chop}_j\text{suffix}_k(w) = a_{n-k+1} \ldots a_{n-j} = \text{suffix}_{k-j}\text{chop}_j(w),$$

and if $n < k$, then $\text{chop}_j\text{suffix}_k(w) = \emptyset = \text{suffix}_{k-j}\text{chop}_j(w)$. Suppose $k < j$. Then $\text{chop}_j\text{suffix}_k(w) = \emptyset = \Phi(w)$.

Consider (c1). Suppose $j \leqslant k$. If $k \leqslant n$, then

$$\text{prefix}_j\text{prefix}_k(w) = a_1 \ldots a_j = \text{prefix}_j\text{chop}_{k-j}(w),$$

and if $n < k$, then $\text{prefix}_j\text{prefix}_k(w) = \emptyset = \text{prefix}_j\text{chop}_{k-j}(w)$. Suppose $k < j$. Then $\text{prefix}_j\text{prefix}_k(w) = \emptyset = \Phi(w)$.

Finally, consider (c2). Suppose $j \leqslant k$. If $k \leqslant n$, then

$$\text{prefix}_j\text{suffix}_k(w) = a_{n-k+1} \ldots a_{n-k+j} = \text{suffix}_j\text{chop}_{k-j}(w),$$

and if $n < k$, then $\text{prefix}_j\text{suffix}_k(w) = \emptyset = \text{suffix}_j\text{chop}_{k-j}(w)$. Suppose $k < j$. Then $\text{prefix}_j\text{suffix}_k(w) = \emptyset = \Phi(w)$. $\square$

We now show that each function in F^+ has at least one canonical form.

Procedure 2.2. Let f be in F^+, say $f = f_1 \ldots f_q$, each f_j in F. Clearly, $f = f\text{chop}^0\text{chop}_0$.

Step 1: Using (a1), (a2) and (a3) repeatedly, move all functions of the form chop_i in the expression $f_1 \ldots f_q\text{chop}^0\text{chop}_0$ as far right as possible. Either Φ arises, in which case the canonical form is $\text{prefix}_2\text{suffix}_1$ and the process terminates, or else use (a4) repeatedly to obtain an expression for f of the form

$$g_1 \ldots g_r\text{chop}_{i_1}. \tag{1}$$

Note that no g_j is of the form chop_i, $r < q+2$, and g_r is of the form chop^j.

Step 2: Using (b1) and (b2) repeatedly, move all functions of the form chop^i in expression (1) as far right as possible. Either Φ arises, in which case the canonical form is $\text{prefix}_2\text{suffix}_1$ and the process terminates, or else use (b3) repeatedly to obtain an expression for f of the form

$$h_1 \ldots h_s\text{chop}^{i_2}\text{chop}_{i_1}. \tag{2}$$

Note that each h_j is either of the form prefix_i or suffix_i, and $0 \leqslant s < r$.

Step 3: If $s = 0$, then $f = \text{chop}^{i_2}\text{chop}_{i_1}$ and the procedure terminates.

Step 4: Suppose $s > 0$. If $s = 1$, then expression (2) is of the form (i) or (ii) in Theorem 1.1 and the procedure terminates. Suppose $s > 1$. Then $h_{s-1}h_s$ is either of the form $\text{prefix}_j\text{prefix}_k$, $\text{prefix}_j\text{suffix}_k$, $\text{suffix}_j\text{suffix}_k$ or $\text{suffix}_j\text{prefix}_k$. If $j > k$, then the canonical form for f is $\text{prefix}_2\text{suffix}_1$ and the process terminates. If $j \leqslant k$, use one application of (c1), (c2), (d1) or (d2) to change $h_{s-1}h_s$ to an expression of the form $\text{prefix}_j\text{chop}_{k-j}$, $\text{suffix}_j\text{chop}_{k-j}$, $\text{suffix}_j\text{chop}^{k-j}$ or $\text{prefix}_j\text{chop}^{k-j}$.

Step 4a: If $h_{s-1}h_s$ is converted to the first or second form, repeat Step 1 on expression (2), obtaining an expression for f either

(i) of the form prefix_2 suffix_1 and halting or

(ii) of the form

$$h'_1 \ldots h'_{s-1}\text{chop}^{i'_2}\text{chop}_{i'_1} \tag{3}$$

and repeat Step 4 with $h_1 \ldots h_s\text{chop}^{i_2}\text{chop}_{i_1}$ replaced by expression (3).

Step 4b: If $h_{s-1}h_s$ is converted to the third or fourth form, repeat Step 2 on expression (2), obtaining an expression for f either

(i) of the form $\text{prefix}_2\ \text{suffix}_1$ and halting or

(ii) of the form

$$h'_1 \ldots h'_{s-1}\text{chop}^{i'_2}\text{chop}_{i'_1} \tag{4}$$

and repeat Step 4 with $h_1 \ldots h_s\text{chop}^{i_2}\text{chop}_{i_1}$ replaced by expression (4).

Note that Procedure 2.2 terminates since Step 4 is used at most (the original) s times. The end result is a canonical expression for f.

Summarizing, we have the following theorem.

Theorem 2.3. *Procedure* 2.2 *yields a canonical expression for each f in F^+.*

Given f in F^+, $f = \Phi$ iff Procedure 2.2 returns the canonical expression prefix_2 suffix_1. In the same spirit, $f = E$ iff applying Procedure 2.2 to f results in an expression of the form $\text{prefix}_0\text{chop}^l\text{chop}_m$ or $\text{suffix}_0\text{chop}^l\text{chop}_m$.

The following example illustrates the use of Procedure 2.2 to obtain a canonical form for a given interval function in F^+.

Example 2.4. Let $f = \text{chop}^2\text{prefix}_5\text{chop}^2\text{suffix}_{10}\text{prefix}_{17}\text{chop}_8$. Then

$$\begin{aligned}
f &= \text{prefix}_3\text{chop}^2\text{chop}^2\text{suffix}_{10}\text{prefix}_{17}\text{chop}_8 && \text{by (b2).}\\
&= \text{prefix}_3\text{suffix}_6\text{chop}^2\text{chop}^2\text{prefix}_{17}\text{chop}_8 && \text{by (b1) twice}\\
&= \text{prefix}_3\text{suffix}_6\text{prefix}_{13}\text{chop}^2\text{chop}^2\text{chop}_8 && \text{by (b2) twice}\\
&= \text{prefix}_3\text{suffix}_6\text{prefix}_{13}\text{chop}^4\text{chop}_8 && \text{by (b3)}\\
&= \text{prefix}_3\text{prefix}_6\text{chop}^7\text{chop}^4\text{chop}_8 && \text{by (d2)}\\
&= \text{prefix}_3\text{prefix}_6\text{chop}^{11}\text{chop}_8 && \text{by (b3)}\\
&= \text{prefix}_3\text{chop}_3\text{chop}^{11}\text{chop}_8 && \text{by (c1)}\\
&= \text{prefix}_3\text{chop}^{11}\text{chop}_{11} && \text{by (a3) and (a4).}
\end{aligned}$$

We now turn to the uniqueness component of Theorem 1.1. To establish this, we need the following auxiliary result.

Lemma 2.5. *Let* $f_1 = \text{prefix}_{j_1}\text{chop}^{k_1}\text{chop}_{l_1}$, $f_2 = \text{suffix}_{j_2}\text{chop}^{k_2}\text{chop}_{l_2}$ *and* $f_3 = \text{chop}^{k_3}\text{chop}_{l_3}$, *where j_1 and j_2 are positive integers and k_i and l_i are nonnegative integers,* $1 \leq i \leq 3$. *Then f_1, f_2 and f_3 are different functions.*

Proof. Let a and b be distinct elements of Σ. Since Σ is assumed to have at least two elements, a and b exist.

Consider f_1 and f_2. Suppose $k_1 \neq k_2$, say $k_1 < k_2$. (A similar argument holds if $k_2 < k_1$.) Then $f_1(a^{k_1}b^{j_1}a^{l_1}) = b^{j_1}$ but (since j_1 is positive) $f_2(a^{k_1}b^{j_1}a^{l_1})$ has less than j_1 occurrences of b. Thus $f_1 \neq f_2$. By symmetry, $f_1 \neq f_2$ if $l_1 \neq l_2$. Suppose $k_1 = k_2$ and $l_1 = l_2$. Then $f_1(a^{k_1}b^{j_1}a^{j_2}b^{l_1}) = b^{j_1}$ but $f_2(a^{k_1}b^{j_1}a^{j_2}b^{l_1}) = a^{j_2}$. Thus $f_1 \neq f_2$.

Now consider f_1 and f_3. But $f_1(a^{k_1+j_1}b^{k_3+l_3+l_1}) = a^{j_1}$ and $f_3(a^{k_1+j_1}b^{k_3+l_3+l_1})$ has at least one occurrence of b. Thus $f_1 \neq f_3$.

By an analogous argument, $f_2 \neq f_3$. □

We are now ready to prove our main theorem.

Proof of Theorem 1.1. By Theorem 2.3, f has a canonical form. By the second paragraph after the statement of Theorem 1.1, the last phrase in the last sentence of the statement of Theorem 1.1 holds. If $f = \Phi$, then f obviously cannot be represented by a form of type (i), (ii), or (iii). Suppose f is in $F^+ - \{\Phi, E\}$. Let

$$F_{\text{pre}} = \{\text{prefix}_k \text{chop}^l \text{chop}_m \mid k \geq 1, l \geq 0, m \geq 0\},$$

$$F_{\text{suf}} = \{\text{suffix}_k \text{chop}^l \text{chop}_m \mid k \geq 1, l \geq 0, m \geq 0\}, \text{ and}$$

$$F_{\text{chop}} = \{\text{chop}^l \text{chop}_m \mid l \geq 0, m \geq 0\}.$$

By Theorem 2.3, f is in $F_{\text{pre}} \cup F_{\text{suf}} \cup F_{\text{chop}}$. By Lemma 2.5,

$$F_{\text{pre}} \cap F_{\text{suf}} = F_{\text{pre}} \cap F_{\text{chop}} = F_{\text{suf}} \cap F_{\text{chop}} = \emptyset.$$

Thus, exactly one of the following cases holds:

(α) f is in F_{pre}.
(β) f is in F_{suf}.
(γ) f is in F_{chop}.

We only give the argument for (α), that for (β) and (γ) being similar.

Consider (α). Suppose f has the canonical forms

$$f' = \text{prefix}_{k_1} \text{chop}^{l_1} \text{chop}_{m_1} \quad \text{and} \quad f'' = \text{prefix}_{k_2} \text{chop}^{l_2} \text{chop}_{m_2}.$$

Let $w_1 = a^{k_1+k_2+l_1+l_2+m_1+m_2}$. Then $a^{k_1} = f'(w_1) = f(w_1) = f''(w_1) = a^{k_2}$. Hence, $k_1 = k_2$. Suppose $l_2 < l_1$. Let $w_2 = a^{l_1}b^{k_1+k_2+l_2+m_1+m_2}$. Then $f'(w_2) = b^{k_1}$ and $f''(w_2)$ has at least one occurrence of a, a contradiction. Similarly, $l_1 < l_2$ leads to a contradiction. Thus, $l_1 = l_2$. Suppose $m_2 < m_1$. Then $f'(a^{k_1+l_1+m_1-1}) = \emptyset \neq a^{k_1} = f''(a^{k_1+l_1+m_1-1})$ since $k_1 \geq 1$, a contradiction. Similarly, $m_1 < m_2$ leads to a contradiction. Thus $m_1 = m_2$, so the canonical forms are the same. □

Using Theorem 2.3 we obtain another set of canonical forms for the functions in F^+. Specifically, we have the following corollary.

Corollary 2.6. *Each function f in F^+ is expressible in the form*

(i) $\mathrm{chop}_m\mathrm{chop}^l\mathrm{prefix}_j$,
(ii) $\mathrm{chop}_m\mathrm{chop}^l\mathrm{suffix}_k$,
(iii) $\mathrm{chop}_m\mathrm{chop}^l$ *or*
(iv) $\mathrm{prefix}_2\mathrm{suffix}_1$.

In addition, if $f \neq E$, then f is in exactly one of the above forms, is uniquely expressible within that form, and if (i) (*respectively* (ii)), *then j* (*respectively k*) *is positive.*

Proof. By methods similar to that of Lemma 2.1, it is easily seen that the following hold for nonnegative integers j and k:

(1) $\mathrm{prefix}_k\mathrm{chop}_j = \mathrm{chop}_j\mathrm{prefix}_{k+j}$,
(2) $\mathrm{suffix}_k\mathrm{chop}_j = \mathrm{chop}_j\mathrm{suffix}_{k+j}$,
(3) $\mathrm{suffix}_k\mathrm{chop}^j = \mathrm{chop}^j\mathrm{suffix}_{k+j}$, and
(4) $\mathrm{prefix}_k\mathrm{chop}^j = \mathrm{chop}^j\mathrm{prefix}_{k+j}$.

By Theorem 1.1, f is in one of the forms

(i′) $\mathrm{prefix}_j\mathrm{chop}^l\mathrm{chop}_m = \mathrm{chop}_m\mathrm{chop}^l\mathrm{prefix}_{j+l+m}$, by Lemma 2.1(a3) and by (1) and (4);
(ii′) $\mathrm{suffix}_k\mathrm{chop}^l\mathrm{chop}_m = \mathrm{chop}_m\mathrm{chop}^l\mathrm{suffix}_{k+l+m}$, by Lemma 2.1(a3) and by (2) and (3);
(iii′) $\mathrm{chop}^l\mathrm{chop}_m = \mathrm{chop}_m\mathrm{chop}^l$, by Lemma 2.1 (a3) or
(iv′) $\mathrm{prefix}_2\mathrm{suffix}_1$ $(=\Phi)$.

Hence, f is expressible in the form (i), (ii), (iii) or (iv).

Suppose $f \neq E$. Now (i′) and (ii′) can be reformulated as

(i″) $\mathrm{chop}_m\mathrm{chop}^l\mathrm{prefix}_j = \mathrm{prefix}_{j-l-m}\mathrm{chop}^l\mathrm{chop}_m$ since $\mathrm{chop}_m\mathrm{chop}^l\mathrm{prefix}_j \neq \Phi$ implies that $j \geq l+m$, and
(ii″) $\mathrm{chop}_m\mathrm{chop}^l\mathrm{suffix}_k = \mathrm{suffix}_{k-l-m}\mathrm{chop}^l\mathrm{chop}_m$ since $\mathrm{chop}_m\mathrm{chop}^l\mathrm{suffix}_k \neq \Phi$ implies that $k \geq l+m$.

Suppose f is expressible by more than one of the types (i)–(iv), or is expressible by two different forms within the same type. Then by (i″), (ii″), (iii′) and (iv′), f is expressible by more than one of the types (i)–(iv) of Theorem 1.1 or is expressible by two different forms within the same type of Theorem 1.1, a contradiction.

Finally, observe that the last phrase of the last sentence in the statement of the corollary is obvious. □

Employing Theorem 1.1, we can also prove that there is no set of canonical forms, each of length at most 2, for F^+.

Corollary 2.7. *There is no set $F' \subseteq F^+$ of forms, each of length at most 2, such that every f in F^+ can be expressed by at least one form in F'.*

Proof. Let $f = \mathrm{prefix}_2\mathrm{chop}^2\mathrm{chop}_2$ and $g = g_1g_2$, where g_1 and g_2 are in F. It suffices to show that $f \neq g$.

Since g_1 and g_2 each have four possible forms, g_1g_2 has sixteen possible forms. Eleven of these are on the left-hand side of the equations in Lemma 2.1. For each of these eleven forms, an insertion of either chop_0 (the identity function) or chop^0 (also the identity function) in an appropriate position in the form on the right-hand side of the corresponding equation yields one of the three nontrivial canonical forms of Theorem 1.1. Furthermore, the resulting canonical form differs in some subscript or superscript from f. (For example, suppose $g = \text{prefix}_j\text{suffix}_k$ for some j and k. Since $f \neq \Phi$, $j \leq k$. By Lemma 2.1(c2), $g = \text{suffix}_j\text{chop}_{k-j} = \text{suffix}_j\text{chop}^0\text{chop}_{k-j}$.) Since $f \neq E$, $f \neq g$ by the uniqueness part of Theorem 1.1. The five forms not considered in Lemma 2.1 are $\text{chop}^k\text{chop}_j$, $\text{prefix}_k\text{chop}^j$, $\text{prefix}_k\text{chop}_j$, $\text{suffix}_k\text{chop}^j$ and $\text{suffix}_k\text{chop}_j$. By the uniqueness part of Theorem 1.1, f cannot be of the form $\text{chop}^k\text{chop}_j$. By an appropriate insertion of either chop^0 or chop_0, it is readily seen that f cannot be one of the remaining four forms. □

3. Fraction-interval functions

The previous two sections discussed prefix and suffix operations defined in terms of fixed length, e.g., prefix_k and chop^k (the 'complement' of prefix_k). In this section we consider prefix and suffix operations defined in terms of returning a fixed *fraction* of the input word. These functions arise in a very natural manner, as for example, in the query: 'Retrieve that part of Smith's checking account which runs from the 56% point to the 60% point'.

Formally, the new operations are described as follows.

Notation. For each λ, $0 \leq \lambda \leq 1$, let Fprefix_λ and Fsuffix_λ (the *fraction-interval* functions) be the mappings from Σ^* to Σ^* defined by[6] $\text{Fprefix}_\lambda(a_1 \dots a_n) = a_1 \dots a_{\lfloor \lambda n \rfloor}$ and $\text{Fsuffix}_\lambda(a_1 \dots a_n) = a_{n-\lfloor \lambda n \rfloor+1} \dots a_n$ for each $a_1 \dots a_n$, $n \geq 0$.

Clearly, $\text{Fprefix}_1 = \text{Fsuffix}_1$ and is the identity function. Also, $\text{Fprefix}_0 = \text{Fsuffix}_0 = E$.

Using the above symbolism, the query 'Retrieve that part of Smith's checking account which runs from the 56% point to the 60% point' can be written as $\text{Fsuffix}_{1/12}\text{Fprefix}_{3/5}$. (Indeed, for $w = a_1 \dots a_{100}$, $\text{Fsuffix}_{1/12}\text{Fprefix}_{3/5}(w) = \text{Fsuffix}_{1/12}(a_1 \dots a_{60}) = a_{56} \dots a_{60}$.)

Since λn is nonintegral in general, some rounding off is necessary. Instead of the round down of above, one could round up to the nearest integer. In any case, rounding leads to approximation. For example,

$$a_1 \dots a_n = \text{Fprefix}_\lambda(a_1 \dots a_n)\text{Fsuffix}_{(1-\lambda)}(a_1 \dots a_n)$$

[6] For each real number x, $\lfloor x \rfloor$ is the largest integer not exceeding x, i.e., $\lfloor x \rfloor$ is the integer r such that $r \leq x < r+1$.

if λ is an integer, and

$$(\mathrm{Fprefix}_\lambda(a_1 \dots a_n))(a_{\lfloor \lambda n \rfloor+1})\mathrm{Fsuffix}_{(1-\lambda)}(a_1 \dots a_n)$$

otherwise. Thus, $\mathrm{Fprefix}_\lambda$ and $\mathrm{Fsuffix}_\lambda$ are not quite true 'complementary' functions.

One could define Fchop_λ and Fchop^λ by $\mathrm{Fchop}_\lambda(a_1 \dots a_n) = a_1 \dots a_{n-\lfloor \lambda n \rfloor}$ and $\mathrm{Fchop}^\lambda(a_1 \dots a_n) = a_{\lfloor \lambda n \rfloor+1} \dots a_n$, for all $a_1 \dots a_n$. It is not clear that the introduction of these functions serves any useful purpose since $\mathrm{Fchop}^\lambda(a_1 \dots a_n)$ is close to $\mathrm{Fsuffix}_{(1-\lambda)}(a_1 \dots a_n)$ and $\mathrm{Fchop}_\lambda(a_1 \dots a_n)$ is close to $\mathrm{Fprefix}_{(1-\lambda)}(a_1 \dots a_n)$.

We are interested in obtaining a set of canonical forms for the composition of the fraction-interval functions alone, as well as in conjunction with the four interval functions discussed in the earlier sections. Unfortunately, as we shall see in Theorem 3.4 below, no such result is possible. To establish Theorem 3.4, we need two lemmas. The first gives an approximation for the composition of two Fprefix functions. The second lemma asserts that if f is the composition of Fprefix functions and $f = g$, with $g = g_1 \dots g_s$, each g_i being one of our six types of interval functions, then each g_i is an Fprefix function.

We need the following.

Notation. For all words w and y and positive integer m, write $w \leqslant_m y$ if w is an interval of y and $|y| - |w| \leqslant m$. Let f and g be interval functions and m a positive integer. Write $f \leqslant_m g$ if, for each word w, $f(w)$ exists implies

(i) $g(w)$ exists and

(ii) $f(w) \leqslant_m g(w)$.

Intuitively, $f \leqslant_m g$ means that $g(w)$ contains all the information in $f(w)$ and at most m additional pieces of data.

Clearly, if $f \leqslant_m g$ and $g \leqslant_n h$, then $f \leqslant_{m+n} h$.

Lemma 3.1. *For all λ and μ, $0 < \lambda, \mu < 1$,* $\mathrm{Fprefix}_\lambda \mathrm{Fprefix}_\mu \leqslant_1 \mathrm{Fprefix}_{\lambda\mu}$.

Proof. Let n be an arbitrary nonnegative integer. Then $\mu n = \lfloor \mu n \rfloor + \varepsilon$ for some ε, $0 \leqslant \varepsilon < 1$. Thus $\lfloor \lambda \mu n \rfloor = \lfloor \lambda \lfloor \mu n \rfloor + \lambda \varepsilon \rfloor$, with $0 \leqslant \lambda\varepsilon < 1$, and

$$\lfloor \lambda \lfloor \mu n \rfloor \rfloor \leqslant \lfloor \lambda \mu n \rfloor \leqslant \lfloor \lambda \lfloor \mu n \rfloor \rfloor + 1.$$

Hence,

$$\begin{aligned}\mathrm{Fprefix}_\lambda \mathrm{Fprefix}_\mu (a_1 \dots a_n) &= a_1 \dots a_{\lfloor \lambda \lfloor \mu n \rfloor \rfloor} \\ &\leqslant_1 a_1 \dots a_{\lfloor \lambda \mu n \rfloor} = \mathrm{Fprefix}_{\lambda\mu}(a_1 \dots a_n). \qquad \square\end{aligned}$$

The $\leqslant_1$ relation in Lemma 3.1 cannot be strengthened to equality. For example,

$$\begin{aligned}\mathrm{Fprefix}_{5.9/10}\mathrm{Fprefix}_{10.9/20}(a_1 \dots a_{20}) &= a_1 \dots a_5 \leqslant_1 a_1 \dots a_6 \\ &= \mathrm{Fprefix}_{(5.9)(10.9)/200}(a_1 \dots a_{20}).\end{aligned}$$

Corollary 3.2. *For all* $0<\lambda_1,\ldots,\lambda_m<1$, $\mathrm{Fprefix}_{\lambda_1}\ldots\mathrm{Fprefix}_{\lambda_m} \leqslant_{m-1} \mathrm{Fprefix}_{\lambda_1\ldots\lambda_m}$.

Notation. Let

$$G=F\cup\{\mathrm{Fprefix}_\lambda, \mathrm{Fsuffix}_\mu \mid 0<\lambda,\ \mu<1\},$$

$$G^+=\{g_1\ldots g_n \mid n\geqslant 1, \text{ each } g_i \text{ in } G\}.$$

Lemma 3.3. *Suppose* $g_1\ldots g_s=\mathrm{Fprefix}_{\lambda_1}\ldots\mathrm{Fprefix}_{\lambda_t}$, *s and t positive integers, each* g_i *in G, no* g_i *the identity function, and* $0<\lambda_j<1$ *for each j. Then, for each i,* $1\leqslant i\leqslant s$, *there exists* μ_i, $0<\mu_i<1$, *such that* $g_i=\mathrm{Fprefix}_{\mu_i}$.

Proof. Let $g=g_1\ldots g_s$ and $f=\mathrm{Fprefix}_{\lambda_1}\ldots\mathrm{Fprefix}_{\lambda_t}$. Since $\mathrm{Fprefix}_\lambda(w)\neq\emptyset$ for all λ and w,

$$f(w)\neq\emptyset \quad \text{for all } w. \tag{5}$$

Let n be an integer such that

$$r=\lfloor\lambda_1\lfloor\lambda_2\lfloor\ldots\lfloor\lambda_t n\rfloor\ldots\rfloor\geqslant 1.$$

Clearly n, thus r, exists. Then

$$f(a^n)=a^r\neq\varepsilon. \tag{6}$$

Assume there exists an integer i, $1\leqslant i\leqslant s$, such that $g_i=\mathrm{prefix}_k$ for some k. Suppose $k>0$. Then $g(a^{k-1})=\emptyset\neq f(a^{k-1})$ by (5), a contradiction. Suppose $k=0$. Then, for each word w, either $g(w)=\emptyset$ or $g(w)=\varepsilon$. Then $g(a^n)\neq f(a^n)$ by (6), a contradiction. Thus,

$$\text{no } g_i \text{ is of the form } \mathrm{prefix}_k. \tag{7}$$

Similarly,

$$\text{no } g_i \text{ is of the form } \mathrm{suffix}_k. \tag{8}$$

Suppose that $g_i=\mathrm{chop}^k$ for some i and k. Then $g(a^{k-1})=\emptyset$. By (5), $f(a^{k-1})\neq\emptyset$, a contradiction. Hence,

$$\text{no } g_i \text{ is of the form } \mathrm{chop}^k. \tag{9}$$

Similarly,

$$\text{no } g_i \text{ is of the form } \mathrm{chop}_k. \tag{10}$$

Suppose that $g_i=\mathrm{Fsuffix}_\lambda$ for some i and λ, $0\leqslant\lambda\leqslant 1$. Without loss of generality, we may assume that

$$i \text{ is the largest such integer.} \tag{11}$$

Since $\mathrm{Fsuffix}_1$ is the identity function, $\lambda\neq 1$ by hypothesis. If $\lambda=0$, then $f(a^n)=\varepsilon$ or $\emptyset$, contradicting (6). Thus $\lambda\neq 0$, i.e., $0<\lambda<1$. By (7)-(11), for each j, $i+1\leqslant j\leqslant s$, there exists a μ_j such that $g_j = \text{Fprefix}_{\mu_j}$. Since g_j is not the identity function (by hypothesis), $\mu_j \neq 1$. And, as above, it is seen that $\mu_j \neq 0$. Thus, $0 < \mu_j < 1$. Then, for each positive integer p, $g_{i+1} \ldots g_s(a^p) = a^{q(p)}$ for some nonnegative integer $q(p)$. Obviously, $\lim_{p\to\infty} q(p) = \infty$. Hence, $\lim_{p\to\infty}(q(p)-1)/q(p) = 1$. Since $0 < \lambda < 1$, there exists a p such that $\lambda < (q(p)-1)/q(p) < 1$. Then,

$$\lfloor q(p)\lambda \rfloor \leq q(p)\lambda < q(p) - 1.$$

Let a and b be distinct symbols. Then

$$\begin{aligned} g(ab^{p-1}) &= g_1 \ldots g_{i-1}\text{Fsuffix}_\lambda g_{i+1} \ldots g_s(ab^{p-1}) \\ &= g_1 \ldots g_{i-1}\text{Fsuffix}_\lambda(ab^{q(p)-1}) \quad \text{since } g_{i+1} \ldots g_s(a^p) = a^{q(p)} \\ &= g_1 \ldots g_{i-1}(b^{\lfloor q(p)\lambda \rfloor}) \quad \text{since } \lfloor q(p)\lambda \rfloor < q(p) - 1. \end{aligned}$$

Hence, $f(ab^{q(p)-1}) \neq g(ab^{q(p)-1})$, since a does not occur in $g(ab^{p-1})$ but does in $f(ab^{q(p)-1})$. This is a contradiction. Therefore,

$$\text{no } g_i \text{ is of the form Fsuffix}_\lambda, \quad 0 \leq \lambda \leq 1. \tag{12}$$

By (7)-(10) and (12), each g_j is of the form Fprefix_μ, $0 \leq \mu \leq 1$. As above, no g_j is of the form Fprefix_0 or Fprefix_1. Hence the lemma holds. □

We are now ready for our result on the nonexistence of a set of canonical forms for G^+, as well as for the composition of the fraction-interval functions alone.

Theorem 3.4. *There is no positive integer t such that each composition of* Fprefix *functions is the composition of at most t functions in G. Thus, there is no positive integer t such that each composition of fraction-interval functions alone, as well as with the other interval functions in G, is the composition of at most t functions in G.*

Proof. It suffices to verify the first sentence only. The idea of the proof is as follows: Given t, a function f of the form $\text{Fprefix}_{\lambda_1} \ldots \text{Fprefix}_{\lambda_{t+1}}$ is chosen so that the evaluation of f on a^{t+1} involves $t+1$ occurrences of round off, with $f(a^{t+1}) = \varepsilon$. However, if $f = g_1 \ldots g_s$, $s \leq t$ and each g_i in G, then there are at most s occurrences of round off, with $g_1 \ldots g_s(a^{t+1}) \neq \varepsilon$.

Formally, let t be a given positive integer. Let δ be a real number, $0 < \delta < 1$, such that

$$(1-\delta)(1-\delta/2) \cdots (1-\delta/(t+1)) > t/(t+1). \tag{13}$$

Since

$$\lim_{\delta\to 0}(1-\delta)(1-\delta/2) \ldots (1-\delta/(t+1)) = 1,$$

such a δ exists. Let

$$f = \text{Fprefix}_{\lambda_1} \ldots \text{Fprefix}_{\lambda_{t+1}},$$

where $\lambda_k = 1 - \delta/k$ for each k, $1 \leq k \leq t+1$. Note that f is not the identity function. (Indeed, let n be an integer such that $r = \lfloor \lambda_1 \lfloor \lambda_2 \ldots \lfloor \lambda_{t+1} n \rfloor \ldots \rfloor \geq 1$. Then $f(a^n) = a^r$, but $r < n$ since $\lambda_i < 1$ for all i.) Let $g = g_1 \ldots g_s$, $s \leq t$, each g_j in G. To complete the proof it is enough to show that $f \neq g$.

Suppose $f = g$. Since f is not the identity function, we may assume without loss of generality that no g_j is the identity function. By Lemma 3.3, for each j, $1 \leq j \leq s$, there exists μ_j, $0 < \mu_j < 1$, such that $g_j = \mathrm{Fprefix}_{\mu_j}$. Note the following easily established result:

(∗) *Let h be a function from the nonnegative integers to the nonnegative integers such that $\lim_{m\to\infty} h(m)/m$ exists. Then, for each real number λ, $\lim_{m\to\infty} \lfloor \lambda h(m) \rfloor / m$ exists and is*[7] $\lambda \lim_{m\to\infty} \lfloor h(m) \rfloor / m$.

For each positive integer m, let

$$m_1 = \lfloor \lambda_1 \lfloor \ldots \lfloor \lambda_{t+1} m \rfloor \ldots \rfloor \quad \text{and} \quad m_2 = \lfloor \mu_1 \lfloor \ldots \lfloor \mu_s m \rfloor \ldots \rfloor.$$

Then $a^{m_1} = f(a^m) = g(a^m) = a^{m_2}$. Thus $m_1 = m_2$. Hence,

$$\begin{aligned}
\lambda_1 \ldots \lambda_{t+1} &= \lim_{m\to\infty} \frac{\lfloor \lambda_1 \lfloor \ldots \lfloor \lambda_{t+1} m \rfloor \ldots \rfloor}{m} && \text{by repeated use of } (*) \\
&= \lim_{m\to\infty} \frac{m_1}{m} \\
&= \lim_{m\to\infty} \frac{m_2}{m} \\
&= \lim_{m\to\infty} \frac{\lfloor \mu_1 \lfloor \mu_2 \lfloor \ldots \lfloor \mu_s m \rfloor \ldots \rfloor}{m} \\
&= \mu_1 \ldots \mu_s && \text{by repeated use of } (*),
\end{aligned}$$

that is,

$$\lambda_1 \ldots \lambda_{t+1} = \mu_1 \ldots \mu_s. \tag{14}$$

Let $\lambda = \lambda_1 \ldots \lambda_{t+1}$ $(= \mu_1 \ldots \mu_s)$. Recall that $\lambda_i = 1 - (\delta/i)$ for each i. Clearly,

$$\mathrm{Fprefix}_{\lambda_i}(a^i) = \mathrm{Fprefix}_{(i-\delta)/i}(a^i) = a^{i-1}$$

for each i, $1 \leq i \leq t+1$. Hence,

$$f(a^{t+1}) = \mathrm{Fprefix}_{\lambda_1} \ldots \mathrm{Fprefix}_{\lambda_{t+1}}(a^{t+1}) = a^0 = \varepsilon. \tag{15}$$

We shall show that $g(a^{t+1}) \neq \varepsilon$, thereby effecting a contradiction and establishing the theorem.

[7] Indeed, for $\lambda \geq 0$, $\lambda \lim_{m\to\infty} \lfloor h(m) \rfloor / m = \lim_{m\to\infty} (\lambda \lfloor h(m) \rfloor - 1)/m \leq \underline{\lim}_{m\to\infty} \lfloor \lambda h(m) \rfloor / m \leq \overline{\lim}_{m\to\infty} \lfloor \lambda h(m) \rfloor / m \leq \lim_{m\to\infty} (\lambda \lfloor h(m) \rfloor + \lambda)/m = \lambda \lim_{m\to\infty} \lfloor h(m) \rfloor / m$. An analogous argument holds for $\lambda < 0$.

By (13), $t/(t+1) < (1-\delta)\cdots(1-\delta/(t+1)) = \lambda_1\ldots\lambda_{t+1} = \lambda$. Thus $t < \lambda(t+1)$, so $s \leq t < \lambda(t+1)$. Then

$$s-1 < s \leq \lfloor \lambda(t+1) \rfloor. \tag{16}$$

Now

$$\begin{aligned} g &\leq_{s-1} \text{Fprefix}_{\mu_1\ldots\mu_s}, && \text{by Corollary 3.2} \\ &= \text{Fprefix}_\lambda, && \text{by (14).} \end{aligned}$$

Hence, $g(a^{t+1}) \leq_{s-1} \text{Fprefix}_\lambda(a^{t+1}) = a^{\lfloor \lambda(t+1) \rfloor}$. Then,

$$\begin{aligned} |g(a^{t+1})| &\geq \lfloor \lambda(t+1) \rfloor - (s-1) && \text{by definition of } \leq_{s-1} \\ &> 0 && \text{by (16).} \end{aligned}$$

Therefore, $g(a^{t+1}) \neq \varepsilon$, as was desired. □

We conclude with some discussion and problems for further investigation.

(1) In view of Theorem 3.1, there is no set of canonical forms for G^+. Is it possible to find a set of canonical forms which 'approximate' the functions in G^+? One reasonable notion for approximation is the $\leq_s$ relation already defined. Do there exist s and t with the following property: For each function $f = f_1 \ldots f_m$ in G^+ there is a function g in $G' = \{g_1 \ldots g_t \mid \text{each } g_i \text{ in } H\}$ such that $f \leq_s g$? Probably not. How about if s is a linear function of m? That is, does there exist an integer t and linear function s from the integers to the integers with the following property: For each positive integer m and function $f = f_1 \ldots f_m$, each f_i in G, there exists a g in G' such that $f \leq_{s(m)} g$?

(2) A second problem area concerns equality. Can one determine if two functions f and g in G^+ are equal? Closely related to this is the question: 'Is there a finite sound and complete set of inference or replacement 'rules' for transforming expressions of G^+ into equivalent expressions in G^+?' For example, one can show that the following rules are sound (for all λ and k, k a nonnegative integer):

(α) $\text{Fprefix}_\lambda \text{prefix}_k = \text{prefix}_{\lfloor \lambda k \rfloor} \text{prefix}_k$,
(β) $\text{Fsuffix}_\lambda \text{suffix}_k = \text{suffix}_{\lfloor \lambda k \rfloor} \text{suffix}_k$,
(γ) $\text{Fprefix}_\lambda \text{suffix}_k = \text{prefix}_{\lfloor \lambda k \rfloor} \text{suffix}_k$, and
(δ) $\text{Fsuffix}_\lambda \text{prefix}_k = \text{suffix}_{\lfloor \lambda k \rfloor} \text{prefix}_k$.

Is this set of rules complete, i.e., for arbitrary f and g in G^+, does $f = g$ imply that f can be transformed into g by repeated application of the equalities of type (α)–(δ) and those in Lemma 2.1? Probably not.

References

1 S. Ginsburg and K. Tanaka, Computation-tuple sequences and object histories, *ACM Trans. Database Systems* 39 (1985) 309–317.
2 S. Ginsburg and K. Tanaka, Interval queries on object histories, Technical Report TR84302, Computer Science Department, Univ. of Southern California, 1984.